

11-1-2016

Bayesian Inference for Median of the Lognormal Distribution

K. Aruna Rao

SDM Degree College, Ujire, India, arunaraomu@gmail.com

Juliet Gratia D'Cunha

Mangalore University, Mangalagangothri, India, gratiajuliet@gmail.com

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Rao, K. Aruna and D'Cunha, Juliet Gratia (2016) "Bayesian Inference for Median of the Lognormal Distribution," *Journal of Modern Applied Statistical Methods*: Vol. 15 : Iss. 2 , Article 32.

DOI: 10.22237/jmasm/1478003400

Bayesian Inference for Median of the Lognormal Distribution

Cover Page Footnote

Acknowledgements The second author would like to thank Government of India, Ministry of Science and Technology, Department of Science and Technology, New Delhi, for sponsoring her with an INSPIRE fellowship, which enables her to carry out the research program which she has undertaken. She is much honored to be the recipient of this award.

Bayesian Inference for Median of the Lognormal Distribution

K. Aruna Rao
SDM Degree College (Autonomous)
Ujire, India

Juliet G. D’Cunha
Mangalore University
Mangalagangorthri, India

Lognormal distribution has many applications. The past research papers concentrated on the estimation of the mean of this distribution. This paper develops credible interval for the median of the lognormal distribution. The estimated coverage probability and average length of the credible interval is compared with the confidence interval using Monte Carlo simulation.

Keywords: Lognormal distribution, credible interval, coverage probability, confidence interval, Monte Carlo simulation

Introduction

The lognormal distribution is widely used in the analysis of rainfall (Ananthakrishnan, & Soman, 1989), survival analysis (Kalbfleisch & Prentice, 2002; Lawless, 2003) and in the analysis of stock market data (D’Cunha & Rao, 2014a). Length biased lognormal distribution is used in the analysis of data from oil field exploration studies (Ratnaparkhi & Naik-Nimbalkar, 2012; see reference therein). In the analysis of stock market data, although lognormal distribution is not directly used, analysis is carried out using log transformation which in turn implies that the underlying distribution is lognormal.

The lognormal distribution belongs to log location scale family. The salient feature of the log location scale family is that the coefficient of variation (CV) of the distribution depends only on the scale parameter and not on the location parameter. When the log location scale family of distribution is obtained through the symmetric location scale family, the median of the distribution is $\exp(\text{location parameter})$. In this case the median is invariant under distributional transformation

Dr. Rao is a Visiting Professor in the Department of Statistics. Email him at: arunaraomu@gmail.com. D’Cunha is a Research Scholar student in the Department of Statistics. Email her at: gratiajuliet@gmail.com.

of the data. The estimation of mean, median and variance of the lognormal distribution has a long history.

Zellner (1971) proposed several estimators of the mean and median of the lognormal distribution and obtained minimum mean square error estimator of these parameters in some class of estimators. Bayes estimators of these parameters were also considered. Subsequently, Padgett and Wei (1977) developed Bayes estimator of reliability function for the lognormal distribution. They used two types of priors namely normal prior for mean and gamma prior for the inverse of the scale parameter; the other prior is the vague prior of Jeffrey (see Ghosh, Delampady, & Samanta, 2006).

This was extended by Padgett and Johnson (1983), where they obtained lower bounds on reliability function of the two parameter lognormal distribution. Sarabia, Castillo, Gómez-Déniz, and Vázquez-Polo (2005) proposed a class of bivariate conjugate priors for μ and σ of the lognormal distribution using the conditional specification. Several procedures were also suggested for the estimation of the hyperparameters. Harvey and van der Merwe (2012) compared the Bayesian credible interval for the means and variances of lognormal distribution.

The performance of the credible interval is compared with credible/confidence interval suggested by Zhou and Tu (2000) and Krishnamoorthy and Mathew (2003). In the last section the authors discuss about bivariate lognormal distribution and obtain Bayesian confidence intervals for the difference between two correlated lognormal means and for the ratio of lognormal variances. The conclusion was the Bayes credible interval has shorter length compared to the length of the other intervals.

D'Cunha and Rao (2014a) developed Bayesian credible interval for the CV of the lognormal distribution and compared it with the confidence interval obtained by the maximum likelihood estimator. They showed that, under mild regularity conditions, Bayes estimator for the mean of the lognormal distribution exists. Thus, research after Zellner (1971) did not focus on median of the lognormal distribution.

The lognormal distribution is positively skewed and for skewed distributions, median is a better estimator rather than the mean, which is affected by extreme values. In medical studies, median survival time is often reported than the mean survival time. This motivates deriving Bayes credible interval for the median of the lognormal distribution. Under absolute error loss function, the Bayes estimator is the minimum average risk estimator.

Among the Bayesian significance tests, for testing a sharp null hypotheses namely the test based on credible interval, the Bayes factor and the full Bayesian significance test (FBST), credible interval is the simplest to compute and straight forward decision can be taken as in the case of significance tests. Thus, the purpose of this study is to compare the performance the Bayes credible interval with the confidence interval obtained from the maximum likelihood estimator (MLE).

Bayes Estimator for the Median of the Lognormal Distribution

Let μ and σ denote the log location and scale parameter of the lognormal distribution. Given a random sample of size n , x_1, \dots, x_n from this distribution, let $Z_i = \log X_i$, $i = 1, \dots, n$, where Z follows normal distribution with parameters μ and σ^2 and maximum likelihood estimator of μ and σ^2 are \bar{Z} and S_z^2 respectively,

where $\bar{Z} = \frac{\sum_{i=1}^n Z_i}{n}$ and $S_z^2 = \frac{\sum_{i=1}^n (Z_i - \bar{Z})^2}{n}$. Using invariance property of

maximum likelihood estimator (Kale (1999)), the maximum likelihood estimate (MLE) of the median of the lognormal distribution namely $\theta = e^\mu$ is given by $\hat{\theta} = e^{\hat{\mu}} = e^{\bar{Z}}$. The asymptotic variance of $\hat{\theta}$ can be obtained using delta method and is given by

$$\text{var}(\hat{\theta}) = e^{2\mu} \frac{\sigma^2}{n} + o(n^{-1}) \quad (1)$$

In the above expression $\frac{n-1}{n} \approx 1$. The 100 (1 - α)% asymptotic confidence interval for $\hat{\theta}$ is given by $\hat{\theta} \pm Z_{\alpha/2} S.E.(\hat{\theta})$, where $Z_{\alpha/2}$ refers to upper $\alpha/2$ th percentile value of the standard normal distribution and $S.E.(\hat{\theta})$ refers to estimated standard error ($\hat{\theta}$). The estimate of μ and σ^2 is obtained by substituting the value of $\hat{\mu}$ and $\hat{\sigma}^2$ in the expression for variance of $\hat{\theta}$.

Four objective priors are considered: the uniform prior $\pi(\mu, \sigma) = 1$, right invariant prior $\pi(\mu, \sigma) = \frac{1}{\sigma}$, left invariant Jeffreys prior given by $\pi(\mu, \sigma) = \frac{1}{\sigma^2}$ and Jeffreys rule prior $\pi(\mu, \sigma) = \frac{1}{\sigma^3}$. For a discussion of these priors see Berger (1985) and Ghosh et al. (2006).

The choice of the right invariant prior stems from the fact that Z follows normal distribution and the right invariant prior used in this paper is the one that is suggested for location scale family (Ghosh et al., 2006). The advantage of objective Bayesian analysis is that the prediction remains the same irrespective of the decision maker. The procedure can be applied universally given the past data. The posterior density $\pi(\mu, \sigma | z_1, \dots, z_n)$ for the uniform, right invariant, left invariant Jeffreys priors and Jeffreys rule prior are given by the following expressions,

$$\pi(\mu, \sigma | z_1, \dots, z_n) = \frac{1}{\sqrt{2\pi} \frac{\sigma}{\sqrt{n}}} e^{-\frac{1}{2} \frac{(\bar{z} - \mu)^2}{\sigma^2/n}} \frac{\left(\frac{(n-1)s_z^2}{2}\right)^{\frac{n+1}{2}}}{\Gamma\left(\frac{n+1}{2}\right)} \quad (2)$$

$$\left(\frac{1}{\sigma^2}\right)^{\left(\frac{n+1}{2}\right)-1} e^{-\frac{1}{2}\left(\frac{n-1}{\sigma^2}\right)s_z^2} \quad (\text{using uniform prior})$$

$$\pi(\mu, \sigma | z_1, \dots, z_n) = \frac{1}{\sqrt{2\pi} \frac{\sigma}{\sqrt{n}}} e^{-\frac{1}{2} \frac{(\bar{z} - \mu)^2}{\sigma^2/n}} \frac{\left(\frac{(n-1)s_z^2}{2}\right)^{\frac{n+2}{2}}}{\Gamma\left(\frac{n+2}{2}\right)} \quad (3)$$

$$\left(\frac{1}{\sigma^2}\right)^{\left(\frac{n+2}{2}\right)-1} e^{-\frac{1}{2}\left(\frac{n-1}{\sigma^2}\right)s_z^2} \quad (\text{using right invariate prior})$$

$$\pi(\mu, \sigma | z_1, \dots, z_n) = \frac{1}{\sqrt{2\pi} \frac{\sigma}{\sqrt{n}}} e^{-\frac{1}{2} \frac{(\bar{z} - \mu)^2}{\sigma^2/n}} \frac{\left(\frac{(n-1)s_z^2}{2}\right)^{\frac{n+3}{2}}}{\Gamma\left(\frac{n+3}{2}\right)} \quad (4)$$

$$\left(\frac{1}{\sigma^2}\right)^{\left(\frac{n+3}{2}\right)-1} e^{-\frac{1}{2}\left(\frac{n-1}{\sigma^2}\right)s_z^2} \quad (\text{using left invariate prior})$$

$$\pi(\mu, \sigma | z_1, \dots, z_n) = \frac{1}{\sqrt{2\pi} \frac{\sigma}{\sqrt{n}}} e^{-\frac{1}{2} \frac{(\bar{z} - \mu)^2}{\sigma^2/n}} \frac{\left(\frac{(n-1)s_z^2}{2}\right)^{\frac{n+4}{2}}}{\Gamma\left(\frac{n+4}{2}\right)} \quad (5)$$

$$\left(\frac{1}{\sigma^2}\right)^{\left(\frac{n+4}{2}\right)-1} e^{-\frac{1}{2}\left(\frac{n-1}{\sigma^2}\right)s_z^2} \quad (\text{using Jeffreys rule prior})$$

Although the independent prior is used for μ and σ , the posterior density has a bivariate correlated distribution. The Bayes estimator of θ is $E(\theta | z_1, \dots, z_n)$, where expectation is taken with respect to the posterior density of μ and σ .

Monte Carlo Simulation

Observations of size n are generated from normal distribution with mean μ and variance σ^2 . The Bayes estimator of the median of the lognormal distribution is $E(e^\mu)$, where expectation is taken with respect to the posterior density of $\pi(\mu, \sigma | \text{data})$. This expectation has no closed form solution and Monte Carlo integration is carried out using importance sampling approach. Since the posterior density is the product of $\eta = 1/\sigma^2$ which is gamma and the conditional density of $\mu | \sigma^2$ which is normal, an observation is generated for η from gamma density and using this value of η , an observation is generated for μ from normal density. This constitutes a pair of observations (η, μ) from the bivariate posterior density.

Using 10,000 simulations the Bayes estimator for the median of the lognormal distribution is given by $E(e^\mu | z_1, \dots, z_n) = \frac{1}{M} \sum_{i=1}^M (e^{\mu_i})$, where M denotes the number of paired samples generated from the posterior distribution. In this paper we have used $M = 10,000$. The equitailed credible interval has the limit $(1 - \alpha/2)^{\text{th}}$ and $\alpha/2^{\text{th}}$ upper percentile value of the posterior distribution of e^μ . For each sample the confidence interval is given by $\hat{\theta} \pm Z_{\alpha/2} S.E.(\hat{\theta})$, where $\hat{\theta} = e^{\hat{\mu}}$ and $Z_{\alpha/2}$ refers to upper $\alpha/2^{\text{th}}$ percentile value of the standard normal distribution and $S.E.(\hat{\theta})$ refers to estimated standard error $(\hat{\theta})$. Using 1000 simulations, the proportion of times the true median lies inside the credible/confidence interval gives the estimated 100 $(1 - \alpha)$ % credible/confidence level and is referred as coverage probability for brevity. For the investigation the

value of μ is fixed at $\log(1000)$ and the CV value used for the investigation ranges from 0.1, 0.3, 0.5, 0.7, 1, 1.5, 2 and 2.5. It may be recalled that $CV = (e^{\sigma^2} - 1)^{1/2}$.

The sample sizes considered are $n = 10, 20, 40, 60, 80, 100, 150$ and 200 . The total number of configurations are 256 (8 sample size \times 8 CV values \times 4 priors). The average time required for the computation of credible interval is approximately 30 minutes for each sample size and CV.

Results

Presented in Table 1 are the number of times coverage probability is maintained by the credible/confidence interval for 8 combinations of CV across sample sizes. We say that a credible/confidence interval maintains credible/confidence level of $(1 - \alpha) = 0.95$ if the coverage probability is in the interval of 0.940 to 0.960. The confidence interval maintains the level for a sample size of $n \geq 60$ and the credible interval maintains the level for a sample size $n \geq 80$. The table also presents the average length of the credible/confidence interval. The average length is computed whenever the credible/confidence level is maintained.

Table 1. Coverage probability of the credible and confidence interval for the Median across sample sizes for 8 combinations of specified values of CV

| n | Bayes Procedure (Equitailed) | | | | | | | | MLE(Equitailed) | |
|---------|--|----|----|----|----------------|--------|--------|--------|--|----------------|
| | Number of times coverage probability is maintained | | | | Average Length | | | | Number of times coverage probability is maintained | Average length |
| | U | R | L | JR | U | R | L | JR | | |
| 10 | 0 | 0 | 0 | 0 | * | * | * | * | 0 | * |
| 20 | 3 | 1 | 2 | 0 | 604.93 | 612.17 | * | * | 1 | 86.03 |
| 40 | 4 | 3 | 3 | 1 | 544.27 | 449.52 | 279.71 | 275.94 | 3 | 492.29 |
| 60 | 5 | 7 | 3 | 3 | 433.82 | 489.59 | 485.34 | 481.19 | 7 | 436.37 |
| 80 | 6 | 6 | 6 | 5 | 395.89 | 393.52 | 391.06 | 440.87 | 6 | 296.97 |
| 100 | 7 | 7 | 8 | 6 | 273.59 | 252.46 | 251.19 | 272.75 | 7 | 217.95 |
| 150 | 6 | 5 | 6 | 5 | 305.89 | 304.94 | 244.34 | 273.68 | 5 | 254.19 |
| 200 | 7 | 5 | 7 | 7 | 185.89 | 185.49 | 202.39 | 184.57 | 5 | 197.64 |
| overall | 38 | 34 | 35 | 27 | 392.04 | 383.96 | 309.00 | 321.5 | 34 | 283.06 |

*Note. Whenever coverage probability is not maintained average length has not been calculated. U-Uniform prior, R-Right invariant prior, L-Left invariant prior, JR-Jeffreys rule prior.

Presented in Table 2 are the average length of the credible interval and the confidence interval for all the sample sizes. The table has been constructed such that the average length of the credible interval is computed over the confidence interval where the nominal confidence level is maintained for each of the prior.

BAYESIAN INFERENCE FOR MEDIAN LOGNORMAL DISTRIBUTION

The average length of the credible interval is shorter compared to confidence interval.

Table 2. Average length of the credible/confidence interval for all sample size for each prior.

| Prior | Number of times coverage probability maintained out of 64 configurations | Number of times credible interval has shorter length | Number of times confidence interval has shorter length | Average length of the credible interval | Average length of the confidence interval |
|-----------------|--|--|--|---|---|
| Uniform | 38 | 18 | 20 | 196.77 | 197.32 |
| Right invariant | 35 | 27 | 8 | 243.23 | 244.64 |
| Left invariant | 30 | 30 | 0 | 292.39 | 294.58 |
| Jeffreys Rule | 27 | 27 | 0 | 304.47 | 308.38 |

An investigation was also carried out to find out the influence of the variability in the data regarding coverage probability and length of the credible/confidence interval. Presented in Table 3 are the coverage probability and length of the credible/confidence interval for sample size $n = 100$. The length of the confidence/credible interval increases with the increase in value of CV, upto the value of $CV = 2$, then there is a decrease in the length for the value of $CV = 2.5$, specific reason for this type of behavior is not known. The length of the credible interval for most of the values of CV is marginally shorter than the confidence interval.

Table 3. Length of the confidence/credible interval for various values of CV when sample size = 100.

| Sample size | Conf/cred interval based on | Length(Coverage probability) when CV equal to | | | | | | | |
|-------------|-----------------------------|---|---------|---------|---------|---------|---------|---------|---------|
| | | 0.1 | 0.3 | 0.5 | 0.7 | 1 | 1.5 | 2 | 2.5 |
| 100 | MLE | 39.08 | 115.16 | 184.99 | 247.49 | 326.88 | 427.06 | 500.25 | 184.99 |
| | | (0.952) | (0.952) | (0.942) | (0.946) | (0.945) | (0.944) | (0.937) | (0.942) |
| | Uniform | 38.95 | 114.79 | 184.49 | 247.29 | 327.26 | 428.90 | 503.76 | 184.49 |
| | | (0.951) | (0.962) | (0.945) | (0.941) | (0.951) | (0.951) | (0.941) | (0.945) |
| | Right | 38.74 | 114.26 | 183.71 | 246.09 | 325.42 | 426.47 | 501.28 | 183.71 |
| | | (0.951) | (0.960) | (0.944) | (0.941) | (0.951) | (0.951) | (0.941) | (0.944) |
| | Left | 38.57 | 113.71 | 182.75 | 244.72 | 324.01 | 424.61 | 498.39 | 182.75 |
| | | (0.946) | (0.959) | (0.944) | (0.940) | (0.946) | (0.946) | (0.940) | (0.944) |
| | Jeffreys Rule | 38.38 | 113.10 | 181.75 | 243.72 | 322.40 | 422.49 | 496.37 | 181.75 |
| | | (0.942) | (0.958) | (0.939) | (0.940) | (0.942) | (0.942) | (0.940) | (0.930) |

Illustrative Example

To illustrate the use of Bayes procedure for the median of lognormal distribution, consider the data on 31 consecutive daily Carbon Monoxide measurements (parts per million - ppm) taken by an oil refinery northeast of San Francisco and nine measurements on the same stack taken by the Bay Area Air Quality Management District (BAAQMD). The data are available from <http://lib.stat.emu.edu/DASL/>. The sample is carbon monoxide measurements taken by the Bay Area Air Quality Management District (BAAQMD). The hypothesis of interest is

$$H_0: \text{Median Carbon Monoxide measurement of BAAQMD} = 58.81 \text{ ppm}$$

The value 58.81 ppm is the estimated median value obtained by using the data on 31 consecutive daily Carbon Monoxide measurements taken by oil refinery to the northeast of San Francisco. The results are summarized in Table 4. Notice that the estimated median value = 58.81 ppm does not lie inside any of the credible/confidence interval, thus we conclude that median carbon monoxide measurement of BAAQMD \neq 58.81 ppm.

Table 4. Credible/confidence interval and length of the credible/confidence interval for 4 priors under Bayes and Maximum Likelihood estimation for BAAQMD data.

| Procedure | Prior | Estimate | Credible/confidence interval | Length of the Credible/confidence interval |
|-----------|---------------|----------|------------------------------|--|
| Bayes | Uniform | 20.27 | (10.60,36.03) | 25.43 |
| | Right | 20.23 | (10.97,34.15) | 23.18 |
| | Left | 20.12 | (11.40,33.29) | 21.90 |
| | Jeffreys Rule | 20.08 | (11.09,33.12) | 22.03 |
| MLE | - | 19.36 | (7.03,31.69) | 24.65 |

Conclusion

The performance of the Bayes credible interval was investigated for the median of the lognormal distribution. It has many applications and most of the previous papers propose credible intervals for the mean of the lognormal distribution. The median of the lognormal distribution depends only on the log location parameter and should be the right choice as the measure of location rather than the mean. Lognormal distribution is right skewed and mean of the distribution is a function of log location and log scale parameter. Thus, the mean is very much influenced by the variability in the data when the underlined distribution is lognormal.

BAYESIAN INFERENCE FOR MEDIAN LOGNORMAL DISTRIBUTION

Confidence interval is computationally simple. The present investigation revealed that Bayes credible interval has shorter average length compared to the confidence interval and is recommended.

Note: A program in the MATLAB software version 7.0 for computation of credible interval is available.

References

- Ananthakrishnan, R., & Soman, M. K. (1989). Statistical distribution of daily rainfall and its association with the co-efficient of variation of rainfall series. *International Journal of Climatology*, 9(5), 485-500. doi:10.1002/joc.3370090504
- Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis* (2nd ed.). New York, NY: Springer-Verlag.
- D’Cunha, J. G., & Rao, K. A. (2014a). Bayesian inference for volatility of stock prices. *Journal of Modern Applied Statistical Methods*, 13(2), 493-505. Available at: <http://digitalcommons.wayne.edu/jmasm/vol13/iss2/29>
- D’Cunha, J. G., & Rao, K. A. (2014b). Bayesian inference for mean of the lognormal distribution. *International Journal of Scientific and Research Publications*, 4(10), 195-203.
- Ghosh, J. K., Delampady, M., & Samanta, T. (2007). *An introduction to Bayesian analysis: Theory and methods*. USA: Springer Science & Business Media.
- Harvey, J., & van der Merwe A. J. (2012). Bayesian confidence intervals for means and variances of lognormal and bivariate lognormal distributions. *Journal of Statistical Planning and Inference*, 142(6), 1294-1309. doi: 10.1016/j.jspi.2011.12.006
- Kalbfleisch, J. D., & Prentice, R. L. (2002). *The statistical analysis of failure time data* (2nd ed.). New York, NY: John Wiley & Sons.
- Kale, B. K. (1999). *A first course on parametric inference* (2nd ed.). Narosa.
- Krishnamoorthy, K., & Mathew, T. (2003). Inferences on the means of lognormal distributions using generalized P-values and generalized confidence intervals. *Journal of Statistical Planning and Inference*, 115(1), 103-121. doi: 10.1016/S0378-3758(02)00153-2
- Lawless, J. F. (2003). *Statistical models and methods for lifetime data* (2nd ed.). New York, NY: John Wiley & Sons.

Padgett, W. J., & Johnson, M. P. (1983). Some Bayesian lower bounds on reliability in the lognormal distribution. *Canadian Journal of Statistics*, 11(2), 137-147. doi: 10.2307/3314981

Padgett, W. J., & Wei, L. J. (1977). Bayes estimation of reliability for the two-parameter lognormal distribution. *Communications in Statistics—Theory and Methods*, 6(5), 443-457. doi: 10.1080/03610927708827505

Ratnaparkhi, M. V., & Naik-Nimbalkar, U. V. (2012). The length-biased lognormal distribution and its application in the analysis of data from oil field exploration studies. *Journal of Modern Applied Statistical Methods*, 11(1), 255-260. Available at: <http://digitalcommons.wayne.edu/jmasm/vol11/iss1/22>

Sarabia, J. M., Castillo, E., Gómez-Déniz, E., & Vázquez-Polo, F. J. (2005). A class of conjugate priors for log-normal claims based on conditional specification. *The Journal of Risk and Insurance*, 72(3), 479-495. doi: 10.1111/j.1539-6975.2005.00133.x

Zellner, A. (1971). Bayesian and non-Bayesian analysis of the log-normal distribution and log-normal regression. *Journal of the American Statistical Association*, 66(334), 327-330. doi: 10.1080/01621459.1971.10482263

Zhou, X.-H., Tu, W. (2000). Interval estimation for the ratio in means of log-normally distributed medical costs with zero values. *Computational Statistics and Data Analysis*, 35(2), 201-210. doi: 10.1016/S0167-9473(00)00009-8